NOTES 22: INFERENCE FOR >2 PROPORTIONS

Stat 120 | Fall 2025

Prof Amanda Luby

What happens when we have more than two categories within a variable? (so response is not Yes/No or True/False)

Example: Some people think that children who are the older ones in their class at school naturally perform better in sports and that these children then get more coaching and encouragement. Could that make a difference in who makes it to the professional level in sports? Below is the birth month of 1478 major league players born since 1975, along with the national birth percentage across the same years.

Month	1	2	3	4	5	6	7	8	9	10	11	12
Obs	137	121	116	121	126	114	102	165	134	115	105	122
Birth %	8%	7%	8%	8%	8%	8%	9%	9%	9%	9%	8%	9%
Exp	118.2	103.4	118.2	118.2	118.2	118.2	133.0	133.0	133.0	133.0	118.2	133.0

How did I find the Expected Count?

Let's walk through an appropriate hypothesis test for this data

(1) Hypotheses

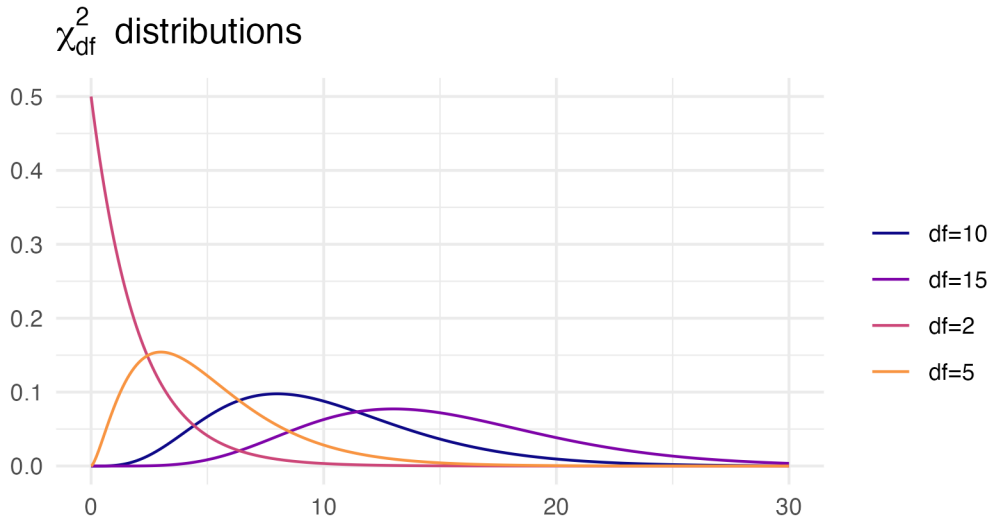
(2) Conditions/Assumptions

(3) Test statistic:

$$\chi^2 = \sum_{\text{all cells}} \frac{(Obs - Exp)^2}{Exp}$$

(4) p-value:

The χ^2 distribution looks like:



For χ^2 tests, we always calculate the p value as

$$P(\chi^2_{n-1} > 26.48) = \texttt{pchisq(26.48, df = 11, lower.tail = FALSE)} = .0055$$

This is a χ^2 Goodness of Fit test.

1 Two Categorical Variables

When we have two categorical variables, each that could have 2 or more levels, the only thing that changes is how we compute the expected counts within each cell.

Example:

	Condition Improved	Did Not Improve	
Drug A			70
Drug B			100
Drug C			150
	180	140	320

(1) Hypotheses

(2) Conditions/Assumptions

(3) Test Statistic

(4) P-value

This is a χ^2 Test for Association

2 Group Problems

1. Below is a sample of sauces. Does Taco Bell distribute sauce packets with equal probability?

2. In order to assess whether habitat conditions are related to the sunlight choices a lizard makes for resting, Western fence lizard (Sceloporus occidentalis) were observed across three different microhabitats. The results are shown below:

	sunlight			
site	sun	partial	shade	Total
desert	16	32	71	119
mountain	56	36	15	107
valley	42	40	24	106
Total	114	108	110	332

a. If the variables describing the habitat and the amount of sunlight are independent, what proportion of lizards (total) would be expected in each of the three sunlight categories?

b. Given the proportions of each sunlight condition, how many lizards of each type would you expect to see in the sun? in the partial sun? in the shade?

	sun	partial	shade	Total
desert				119
mountain				107
valley				106
Total	114	108	110	332

c. Compare the observed (original data) and expected (part b.) tables. From a first glance, does it seem as though the habitat and choice of sunlight may be associated?

d. Perform a χ^2 test to formally test for an association between habitat and choice of sunlight. Write out the appropriate hypotheses, check conditions, compute the test statistic, find the p-value, and interpret your result in context.